July 28, 2020

Dear:

David Landsman: <https://irp.nih.gov/pi/david-landsman>

**Mike Cherry**: <https://profiles.stanford.edu/j-michael-cherry?tab=research-and-scholarship>

Janet Kelso: <https://www.eva.mpg.de/genetics/staff/janet-kelso/index.html?Fsize=0%2CGrawunder%27A%3D0>

Elizabeth Green: OUP, VEO

Simone Larche: Operations

Frances Ouellette: <https://bioinformatics.ca/person/francis-ouellette/>

We are pleased to submit our manuscript titled “Increasing metadata coverage of SRA BioSample entries using deep learning based Named Entity Recognition” by Klie *et al.* for publication in *Database*.

In this manuscript, we address the state of metadata quality in the Sequence Read Archive (SRA) hosted by NCBI. We highlight several issues in the annotations of samples deposited in the SRA, that have similarly been reported for other large data repositories. We propose a novel deep learning-based approach for metadata curation of SRA samples, utilizing Named Entity Recognition (NER) to improve the plurality and completeness of the metadata landscape. This methodology circumvents the need for manual curation, that is often necessary in NER-based efforts of improving metadata quality

The current work is unpublished and is not under consideration elsewhere. All co-authors concur with the contents of the paper. The following individuals would serve as excellent reviewers: The corresponding author’s email address is XXXX.

We look forward to your editorial decision.

Sincerely,